

+ Page 71 +

The Public-Access Computer Systems Review 1, no. 2 (1990):
71-80.

Recursive Reviews

"Public-Access Computer Systems and the Internet"

by Martin Halbert

Recursive Reviews is a new column that will identify and briefly describe articles that deal with public-access computer systems (PACS) and related topics in both library and computer science literature. The "recursive" in the name of the column emphasizes the idea that the discussion of information technology in libraries changes the underlying precepts of the discussion. The dialogue concerning uses of library technology redefines itself in this way, and can therefore be seen as recursive. Enough introduction, let's go on to the reviews.

All followers of the PACS-L forum are aware by now that a great many library catalog systems are accessible via the Internet. The availability of these resources raises a great many questions and possibilities in the library and network user communities. What can be accomplished with this new communications channel? Exactly what is the Internet? What is its extent, and how does it differ from other computer networks? The articles and books reviewed in this column will be of use to anyone having questions about library systems and the Internet, from those unfamiliar with networking technology to those very conversant with it.

+ Page 72 +

Engle, Mary. "Library Systems on the Internet." DLA Bulletin 9,
no. 2 (Fall 1989): 1, 3-4. (ISSN 0272-037X)

Mary Engle's article "Library Systems on the Internet" is a good basic introduction to the recent phenomena of PACS on the Internet. She concisely places the phenomena in the context of advances in library information technology, and she mentions some of the more notable Internet resource experiments that university libraries are undertaking. Basically, libraries have discovered that they can make their catalogs available over the existing Internet computer network, and some libraries are now making many other types of databases and services available. For example,

Engle describes how patrons on the MELVYL system can now directly access the Colorado Alliance of Research Libraries (CARL) collection of library systems. Plans are for the MELVYL catalog to provide access to the RLIN and OCLC bibliographic utilities. Other libraries are mounting periodical indexes. Engle briefly mentions the main problems associated with Internet access to a library's catalog or other resources: incompatibilities between systems accessed via the network, variations in systems' user interfaces and data structures, and increased demands by patrons resulting from the new services.

Larsen, Ronald L. "The Colibratory: The Network as Testbed for a Distributed Electronic Library." Academic Computing 4, no. 5 (February 1990): 22-23, 35-37. (ISSN 0892-4694)

Ronald Larsen speculates in his article "The Colibratory: The Network as Testbed for a Distributed Electronic Library" on the possibilities of experimenting with new library related services on the Internet. He first recounts the history of the national research network from its beginning as the Defense Department's ARPANET in 1969 to the legislation now in progress for the creation of a National Research and Education Network (NREN) for the 1990s. Larsen then outlines EDUCOM president Kenneth King's vision of a world scholarly community that uses the network to communicate electronically and to access collaborative databases via a standardized, intuitive electronic interface. This standard interface would most likely use a network query protocol (a specified format for relaying information) such as NISO Z39.50 to access the many library online public access catalogs and other databases that would be made available (for a criticism of Z39.50 in this role, see Schoffstall's article below, which maintains that Z39.50 requires significant overhaul before it will be useable).

+ Page 73+

Library OPACS are just one of the collaborative information utilities that Larsen envisions as being central to scholarly work of the nineties. Federally produced full-text serials such as the Congressional Record, statistical data like the national census, and other depository information are logical resources for the NREN, since they are public information already. This kind of information can be provided free or for minimal fees, but what about commercially produced databases?

Larsen maintains that the real benefit of the research network vision is in making resources available at little or no direct cost to the user, as books are made available in libraries. This requirement is what has made the realization of an electronic network library so difficult in Larsen's view, and restricted actual experiments to "a small number of pilot projects."

Another significant barrier to users wishing to use the network as a means of accessing information resources is the esoteric nature of today's Internet. Fundamental facts about the network, such as its organization or available information services, are unavailable or hard to find (see the reviews of Comer and Quarterman for the best two tools in this area). Knowledge about how to use the Internet has always been an arcane lore, and network access must become more friendly and understandable before it can be used as a major channel of communication.

Despite these barriers to use, Larsen maintains that the information community must not hesitate to experiment with the Internet as an enabling technological infrastructure. He summarizes: "The concept of a colibratory treats the Internet as a prime environment for collaborative experimentation on high performance distributed information services involving network developers, information resource providers, and network-based consumers."

+ Page 74 +

Kibbey, Mark, and Nancy H. Evans. "The Network is the Library."
EDUCOM Review 24, no.3 (Fall 1989): 15-20. (ISSN 1045-9146)

Mark Kibbey and Nancy Evans have a vision of the future very similar to Larsen's, but focus more on the details of implementation in "The Network is the Library." The frustrations of incompatible systems have undoubtedly been encountered by everyone reading this column. Without standard information formats, retrieval methods, and hardware platforms there can be no economies of scale or common user proficiency in using information systems. Incompatibility of software and hardware is one of the major problems that plague all people who use computer systems, from basic tasks like word processing to the most complex programming. Many libraries are now struggling with the problems of idiosyncratic CD-ROM databases which all have different technical requirements and search interfaces. Kibbey and Evans identify the standards issues that will need to be addressed to avoid similar incompatibility problems when developing information resources on the Internet, echoing many of Larsen's points like the importance of the Z39.50 protocol.

They also discuss the importance of document format standards for bibliographic control and indexing. This is an essential point for the future. We must begin to settle on document formats today for the full text databases that will be built in the future. Formats like the Standardized General Markup Language (SGML) which specify bibliographic information such as author and title should be favored over pure display formats like Postscript, which simply contain page layout formatting. This distinction is important because without labeling bibliographic elements within the text the retrospective process of reformatting an electronic document for a database becomes much

more difficult.

Kibbey and Evans go on to describe Project Mercury, a prototype system that demonstrates all the strengths of the electronic library based on networking standards. Mercury is a full-text indexed electronic library of journal articles, reports, and other current technical literature on artificial intelligence. The project was jointly formed by Carnegie-Mellon University and OCLC to study the possibilities of the new technology. The document format that they used was a proprietary format developed by DEC (also involved in the project) based on the principles emerging from the still-incomplete Office Document Architecture (ODA). ODA is an attempt at achieving the best of both worlds, including both bibliographic element specification like SGML and page layout like Postscript.

+ Page 75 +

Kibbey and Evans conclude, like Larsen, by saying that parties with a "stake in the next generation of academic information services," most particularly librarians, need to aggressively experiment with the new technology and expand their concept of publishing beyond the traditional print view.

Lynch, Clifford A. "Linking Library Automation Systems in the Internet: Functional Requirements, Planning, and Policy Issues." Library Hi Tech 7, no.4 (1989): 7-18. (ISSN 0737-8831)

Clifford Lynch's article could serve as any library administrator's guide to implementation issues when considering involvement with PACS on the Internet. In "Linking Library Automation Systems in the Internet: Functional Requirements, Planning, and Policy Issues" he thoroughly analyzes the practical policy decisions and problems that come up when library systems are accessed over typical network setups. Basic elements such as the operating system one is using on the library system and the types of terminals that one decides to support have significant impacts on how well a system can be used on the network.

Probably the central realization that Lynch offers is that one should study the nitty gritty functional details before getting involved in the network, not after. Ask yourself questions like: Is my system capacity really sufficient to support an indefinite number of users coming in over the network? If I want the system resources prioritized, is my system capable of this? How good is the terminal support on my system? Can it handle a reasonable subset of the constellation of different terminal types out there? Are my security and authentication measures up to keeping out network intrusions? Until one has good answers to all of Lynch's questions, one should stay away from the Internet.

Lynch concludes with a cogent question about integrated library

system vendors. Will these vendors develop complex Internet application features when a relatively small part of their client base is heavily involved with the network? If the vendors do not pursue this technology, will research libraries who wish to move ahead into the networked arena be forced once again into the expensive route of in-house application development?

Have you reached the stage where you are tired of being confused by all the unfamiliar jargon and cryptic acronyms that come up when discussing the Internet? Are you interested in finding out the exact extent of this amorphous creeping electronic vine? Then read on, the next two books are for you.

+ Page 76 +

Comer, Douglas. Internetworking with TCP/IP: Principles, Protocols, and Architecture. Englewood Cliffs, NJ: Prentice-Hall, 1988. (ISBN 0-13-470154-2)

Douglas Comer's book "Internetworking with TCP/IP: Principles, Protocols, and Architecture" is the basic text for anyone seeking to understand the Internet. TCP/IP stands for "Transmission Control Protocol/Internet Protocol." TCP/IP is the fundamental protocol set on which the Internet is built. DARPA converted its ARPANET to this set of protocols between 1980 and 1983, and, in a far-sighted move, funded the implementation of TCP/IP in the UNIX environment, the premiere university computer science operating system. In coordination with the National Science Foundation's new NSFNET and other major government agency networks, the modern day TCP/IP Internet, or just Internet, came into being.

The evolution of the Internet in the last decade has paralleled the general explosion in computer technologies. It was no small feat. In one essentially seamless network, the vast archipelagos of government and research computers are linked, from the largest supercomputers to microcomputers that can sit on our desks. The three core services that make Internet so useful are electronic mail, remote login, and file transfer. Followers of the PACS-L electronic forum are no doubt aware of the advantages of the first two, and may have used FTP (File Transfer Protocol) facilities to retrieve files from the data archives of Internet sites.

The Internet is not yet a commercial product, although some portion of its services will probably become commercial at some future date. Until that time, the Internet will remain an arena for research projects of all kinds (for future technical developments of the Internet, see Mills, et al. below), and it should be investigated by librarians. Comer's book can walk you through all the specifics of the network.

+ Page 77 +

Quarterman, John S. The Matrix: Computer Networks and
Conferencing Systems Worldwide. Bedford, MA: Digital Press, 1990.
(ISBN 1-55558-033-5)

Once you understand the principles behind the Internet, you may want to study the physical layout of the system, and for that you need John Quarterman's book "The Matrix: Computer Networks and Conferencing Systems Worldwide." Quarterman shows how the many world networks interconnect, where their main sites are, and the overall geographic layout of the systems. If I have been overly biased toward the Internet in this column, I apologize. There are many other networks that exist in various degrees of interconnection with Internet, notably BITNET (Because It's Time Network, an academic network similar to Internet, but with lower transmission speeds and no remote login facilities) and UUCP (the name comes from the network's main protocol, Unix to Unix Copy Program, which allows almost anyone with a Unix system and a phone to join the network). There are literally hundreds of networks connecting computer sites worldwide, and if you want to get an overview of them, you have to study Quarterman's book at length (or be a networking guru yourself).

The title Quarterman chose for his book is interesting, and seems to be a case of non-fiction following fiction. Popular science fiction books such as William Gibson's "Neuromancer" were calling the world network system "The Matrix" in the early eighties. It is also interesting to note that Autodesk has chosen to try to implement the interface to the Matrix that Gibson described using the same name, "Cyberspace". How far behind science fiction is today's technology anyway?

+ Page 78 +

Mills, David L., Paul Schragger, and Michael Davis. "Internet Architecture Workshop: Future of the Internet System Architecture and TCP/IP Protocols." Computer Communication Review 20, no.1 (January 1990): 6-17. (ISSN 0146-4833)

We may not have caught up with science fiction, but the Internet is certainly progressing technologically, as described in "Internet Architecture Workshop: Future of the Internet System Architecture and TCP/IP Protocols." There are many exciting new developments coming to the Internet within the time frame of NREN, such as gigabit speeds and millions of additional users. The report of the workshop shows that the people who maintain and develop the Internet are very concerned about the problems involved in keeping the network in good working order as it changes and expands dramatically. David Mills, Chair of the Internet Architecture Task Force, posed a serious question in his session on "Navigation Aids for the Future Internet," saying "We occasionally see cases of Internet routing bobbles, meltdowns and

black holes, even with only 700 nets and uncoordinated back door paths which invite sinister routing loops. Are the Internet addressing and routing algorithms adequate for very large networks with millions of subscribers?" (p. 6)

Some workshop participants questioned the need for a gigabit network (meaning a speed increase of roughly 100 times for the Internet), but were reminded that many scientists needed the high speeds for data transfer. For example, during future space missions researchers hope to engage in global collaboration on the returned data, which will consist of multiple megabyte-size files. At the same time, researchers are concerned that the development of the technical infrastructure of the network not drain critical funding for basic science.

Sound like a familiar library dilemma of trading off automation budgets against other library budget items? There are actually many issues of common interest between Internet workers and librarians. We are both faced with the same dilemma of being pushed toward charging for some services in the future where we did not charge for any in the past. The networking people don't have any magic answers either, but it's instructive to see the same dilemma from another perspective.

+ Page 79 +

Schoffstall, Martin L., and Wengyik Yeong. "A Critique of Z39.50 Based on Implementation Experience." Computer Communication Review 20, no.2 (April 1990): 22-29. (ISSN 0146-4833)

Finally, Martin Schoffstall and Wengyik Yeong give us the benefit of their practical experience in working with the Z39.50 standard in their testbed project on NYSERNet in conjunction with OCLC. The purpose of the Z39.50 Information Retrieval Protocol is standardization of bibliographic queries across networks. Unless a standard of this kind can be agreed upon, we will be facing a chaos of different access methods. Schoffstall and Yeong identify many shortcomings in the current Z39.50, but their points are called into question by their own comments. They complain that the drafters of Z39.50 did not understand the importance of maintaining full compatibility with pre-existing protocols such as the Remote Operations Service (ROS) standard. On the next page they admit that they don't understand the distinctions of the MARC format and actually propose discarding it in favor of a yet-to-be-developed "systematic cataloging method free of redundant specification" (p. 24). Hopefully, the catalogers out there will give these authors the benefit of the developmental history of MARC. They helpfully include their network IDs in their paper: schoff@psi.com and yeongw@nisc.nyser.net

The articles reviewed here are a small subset of the literature on the Internet, but hopefully they will benefit you in studying

the issues involved in implementing library services over the network.

+ Page 80 +

About the Author

Martin Halbert is Automation and Reference Librarian at the Fondren Library of Rice University. He has worked as a corporate librarian and consultant for the IBM corporation. His phone number is (713) 527-8101, extension 2577 and his e-mail address is HALBERT@RICEVM1.RICE.EDU.

The Public-Access Computer Systems Review is an electronic journal. It is sent to participants of the Public-Access Computer Systems Forum, a computer conference on BITNET. To join the PACS Forum, send an electronic mail message to LISTSERV@UHUPVM1 that says:

SUBSCRIBE PACS-L First_Name Last_Name.

The Public-Access Computer Systems Review is copyright (C) 1990 by the University Libraries, University of Houston. All rights reserved.

This article is copyright (C) 1990 by Martin Halbert. All rights reserved.

Copying is permitted for noncommercial use by computerized bulletin board/conference systems, individual scholars, and libraries. Libraries are authorized to add the journal to their collections at no cost. This message must appear on copied material. All commercial use requires permission.
